Radio CovILD, 1-year follow-up

Response to the statistical reviewer

CovILD study team

2022-02-10

# Response to the statistical reviewer

We would like to thank the reviewer for careful study of the manuscript and the feedback. Please find below our response to the open issues.

## Issue 1

**Issue:** In general, the statistical methods and results need pruning. There are too many analyses for this single paper, and aren’t needed for the main aims seem to be focused around the inferential analysis of the risk factors associated with any CT abnormalities (yes/no) and the risk factors associated with CT severity score (0-25). Also, I see the benefit of understanding the trend of CTSS over time.

**Response:** We reduced the amount of data and analyses presented in the manuscript as suggested and followed the specific comments of the editor in the manuscript text. In particular, **Figure 6** comping with the inter-rater comparison of the manual and automated abnormality detection and the plots of effect sizes in **Figure 5B** are removed and the effect size measures cited in the main text instead.

## Issue 2

**Issue:** Methods and results related to the automatic quantification opacity detection seems outside of the scope of this paper. Removing the text and results relating to this would really help the readability of this paper, and make the results clearer.

**Response**: The corresponding text content and **Figure 6** were removed from the revised manuscript.

![Figure 1: Effects of the age variable splining.](data:application/pdf;base64,)

Figure 1: Effects of the age variable splining.

**Figure R1. Effects of the age variable splining.**

## Issue 3

**Issue:** Instead of categorization of the continuous variables, you should considering splining these variables in the regression analyses. Even if the variables are not normal there are better and more powerful ways of handling these variables in the analysis.

**Response:** This is certainly an important point. We agree that manual categorization of numeric variables is controversial and not optimal in terms of model performance. However, categorization of clinical and demographic parameters with widely accepted/published cutoffs is usually accepted better by the clinical community. In our manuscript, we applied cutoffs of BMI and pack-years associated with clinical phenotypes (see: (Pleasants et al., 2020) for smoking). For the age variable, the cutoff choice was motivated by clinical experience of the study team and corresponded approvimately to the mean age of the study cohort.

To address the issue in more detail, we modeled the risk of any CT abnormalities as a function of splined age (logistic GAM, thin plate smoothing method, k = 20, R package *mgcv*, function *gam()*). Interestingly, the inflection point of the predicted risk curve approximates well the chosen age cutoff of 60-years (**Figure R1A**). In general, inclusion of categorized age or smoothed age term in the multi-variate risk modeling (explanatory variables: age, sex and COVID-19 severity) has negligible effect on the accuracy, sensitivity and specificity at predicting any CT abnormalities at the 1-year follow-up (**Figure R1B**).

The comparison of the smoothed and cutoff models are included in the project pipeline deposited at GitHub (<https://github.com/PiotrTymoszuk/radio_CovILD>).

## Issue 4

**Issue:** The univariable models are fine on their own (and I think the associated plots are super nice!), but they should not be used to inform the variables that are included in your multivariable models.

**Response:** We apologize for the unclarity. The multi-parameter models were constructed with the full set of candidate variables (i.e. age, sex, BMI class, smoking quantity, smoking status and acute COVID-19 severity) without any pre-selection based on the results of univariable modeling. The sole criterion for full model term elimination was AIC reduction (function *stepAIC()*, R package *MASS*). We clarify this in the *Statistical analysis* section.

![Figure 2: Comparison of the Poisson and logistic ordinal regression results.](data:application/pdf;base64,)

Figure 2: Comparison of the Poisson and logistic ordinal regression results.

**Figure R2. Comparison of the Poisson and logistic ordinal regression results.**

## Issue 5

**Issue:** Poisson regression doesn’t seem quite right for the CTSS outcome. Ordinal logistic regression would be better since certain outcomes are “higher” or “lower” than others, but you’re not sure by how much or if it even matters how much. In this example, the outcomes are ratings on a 1-25 scale, you know 4 is better than 2 but you can’t necessarily say it’s “twice as good” as 2. Also, a poisson model assumes both an infinite support and a that the mean-variance is the same – both are not true with this outcome.

**Response:** We understand the problem. In the revised manuscript, the severity of CT abnormalities was analyzed by logistic ordinal regression as suggested (function *polr*, R package *MASS*) and the results presented in **Figure E1**. Since, the ordinal models had severe convergence problems while working with the full scale CTSS (0 - 25) and the assumption of odds proportionality (Brant test (Brant, 1990)) was severely violated, we decided to define the following severity classes: 0, 1 - 5, 6 - 10 and 11 - 25 CTSS. A shown in **Figure R2A** for multi-variable modeling of the CT abnormality severity both Poison and ordinal regression returned comparable results, i.e. in both cases age, sex and acute COVID-19 severity were identified as significant predictors. Still, the ordinal model showed only a limited performance at predicting moderate and severe CT abnormalities (see confusion matrix in **Figure R2B**), an issue which may be addressed by a broader set of explanatory variables.

## Issue 6

**Issue:** The word ‘kinetics’ is confusing to me throughout this manuscript. I think you just mean the change or trend of CTTS over time. Given this, I think a repeated measure model is appropriate here. I don’t understand the use of Kendall’s tau the assess this relationship.

**Response:** We have changed the wording throughout the manuscript to ‘change over time’ or ‘rate of improvement’ as suggested by the reviewer and editor. The intention of presenting statistic was to show the reduced drop of CTSS between the consecutive follow-ups, which may be interpreted as decelerated structural lung recovery or chronicity of the residual abnormalities. For the sake of manuscript clarity, we removed the effect size plots from **Figure 5**. Instead, we decided to use more common effect size measure, the non-parametric statistic for paired data (Fritz et al., 2012) and cite its value and effect size interpretation in the main text section describing the changes of CTSS over time.

## Issue 7

**Issue:** The figures are much improved, I think there is just too much presented in each figure. All of the extraneous data in the figure can be removed (i.e. the test statistic, df, etc.).

**Response:** The numeric and text content presented in the Figures was reduced as suggested by the Editor in the manuscript text.

# Other changes introduced to the analysis pipeline

## Change 1

The construction of multi-parameter models by backward elimination was accomplished now with another R tool (function *stepAIC()*, package *MASS* instead of *train()* function from *caret*). With any of the tools, the same set of significant explanatory variables was included in the final multi-parameter models. The OR and 95CI estimates are larger due to change of the baseline level for the sex variable as requested by the editor. See: **Figure 3B**.

# References

Brant R. 1990. Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics* **46**:1171. doi:[10.2307/2532457](https://doi.org/10.2307/2532457)

Fritz CO, Morris PE, Richler JJ. 2012. Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology General* **141**:2–18. doi:[10.1037/A0024338](https://doi.org/10.1037/A0024338)

Pleasants RA, Rivera MP, Tilley SL, Bhatt SP. 2020. Both Duration and Pack-Years of Tobacco Smoking Should Be Used for Clinical Practice and Research. *Annals of the American Thoracic Society* **17**:804–806. doi:[10.1513/ANNALSATS.202002-133VP/SUPPL\_FILE/DISCLOSURES.PDF](https://doi.org/10.1513/ANNALSATS.202002-133VP/SUPPL_FILE/DISCLOSURES.PDF)